

On BICM receivers for TCM transmission

Alex Alvarado, Leszek Szczecinski[§], and Erik Agrell

Department of Signals and Systems, Communication Systems Group

Chalmers University of Technology, Gothenburg, Sweden

[§]INRS-EMT, Montreal, Canada

alex.alvarado@chalmers.se, leszek@emt.inrs.ca, agrell@chalmers.se

Abstract

Recent results have shown that the performance of bit-interleaved coded modulation (BICM) using convolutional codes in nonfading channels can be significantly improved when the interleaver takes a trivial form (BICM-T), i.e., when it does not interleave the bits at all. In this paper, we give a formal explanation for these results and show that BICM-T is in fact the combination of a TCM transmitter and a BICM receiver. To predict the performance of BICM-T, a new type of distance spectrum for convolutional codes is introduced, analytical bounds based on this spectrum are developed, and asymptotic approximations are also presented. It is shown that the minimum distance of the code is not the relevant optimization criterion for BICM-T. Optimal convolutional codes for different constrain lengths are tabulated and asymptotic gains of about 2 dB are obtained. These gains are found to be the same as those obtained by Ungerboeck's one-dimensional trellis coded modulation (1D-TCM), and therefore, in nonfading channels, BICM-T is shown to be asymptotically as good as 1D-TCM.

Index Terms

Bit-interleaved Coded Modulation, Binary Reflected Gray Code, Coded Modulation, Convolutional Codes, Interleaver, Quadrature Amplitude Modulation, Pulse Amplitude Modulation, Set Partitioning, Trellis Coded Modulation.

This work was partially supported by the European Commission under projects NEWCOM++ (216715) and FP7/2007-2013 (236068), and by the Swedish Research Council, Sweden (2006-5599). When this work was submitted for publication, L. Szczecinski was on sabbatical leave with CNRS, Laboratory of Signals and Systems, Gif-sur-Yvette, France. Parts of this work will be presented at the Allerton Conference on Communication, Control, and Computing, IL, USA, September 2010.

I. INTRODUCTION

Coded modulation (CM) was introduced in 1974 when Massey proposed the idea of jointly designing the channel encoder and modulator [1]. This inspired Ungerboeck's trellis coded modulation (TCM) [2], and Imai and Hirakawa's multilevel coding [3]. Bit-interleaved coded modulation (BICM) [4]–[6] appeared in 1992 as an alternative for CM in fading channels. One particularly appealing feature of BICM is that all the operations are bit-wise, i.e., off-the-shelf binary codes and Gray-mapped constellations are used at the transmitter's side and connected via a bit-level interleaver. At the receiver's side, reliability metrics for the coded bits (L-values) are calculated by the demapper, de-interleaved, and then fed to a binary decoder. This structure gives the designer the flexibility to choose the modulator and the encoder independently, which in turn allows, for example, for an easy adaptation of the transmission to the channel conditions (adaptive modulation and coding). This flexibility is arguably the main advantage of BICM over other CM schemes, and also the reason of why it is used in almost all of the current wireless communications standards, e.g., HSPA, IEEE 802.11a/g, IEEE 802.16, and DVB-S2 [6, Ch. 1].

Bit-interleaving before modulation was introduced in Zehavi's original paper [4] on BICM. Bit-interleaving is indeed crucial in fading channels since it guarantees that consecutive coded bits to be sent over symbols affected by independent fades. This results in an increase (compared to TCM) of the so-called code diversity (the suitable performance measure in fading channels), and therefore, BICM is the preferred alternative for CM in fading channels. BICM can also be used in nonfading channels. However, in this scenario, and compared with TCM, BICM gives a smaller minimum Euclidean distance (the proper performance metric in nonfading channels), and also a smaller constraint capacity [5]. If a Gray labeling is used, the capacity loss is small, and therefore, BICM is still considered valid option for CM over nonfading channels. However, the decrease in minimum Euclidean distance makes BICM less appealing than TCM in nonfading channels.

The use of a bit-level interleaver in nonfading channels has been inherited from the original works on BICM by Zehavi [4] and Caire *et al.* [5]. It simplifies the performance analysis of BICM and is implicitly considered mandatory in the literature. However, the reasons for its presence are seldom discussed.

Previously, we have shown in [7] how—by using multiple interleavers—the performance

of BICM can be improved in nonfading channels. Recently, however, it has been shown in [8] that in nonfading channels, considerably larger gains (a few decibels) can be obtained if the interleaver is *completely removed* from the transceiver's configurations. In other words, it was shown that in nonfading channels BICM without an interleaver performs better than the conventional configurations of [4], [5]. The results presented in [8] are solely numerical and an explanation behind such an improvement is not given. In particular, [8] does not explain why the obtained gains depend on the constraint length of the convolutional code (CC). Nevertheless, in [8] some intuitive explanations (using the notion of unequal error protection) and a bit labeling optimization are presented.

In this paper, we present a formal study of BICM with trivial interleavers (BICM-T) in nonfading channels, i.e., the BICM system introduced in [8] where no interleaving is performed. We recognize BICM-T as the combination of a TCM transmitter and a BICM receiver and we develop analytical bounds that give a formal explanation of why BICM-T with CCs performs well in nonfading channels. We also introduce a new type of distance spectrum for the CCs which allows us to analytically corroborate the results presented in [8]. These gains are shown to appear even for one of the simplest configuration one could think of, i.e., when the constraint length $K = 3$ convolutional code with generators $(5, 7)$ is used together with 4-ary pulse amplitude modulation (PAM). Asymptotic bounds are also developed and used to show that for the $(5, 7)$ code and 4-PAM, an asymptotic gain of 2.55 dB is obtained compared to an uncoded system with the same spectral efficiency. Motivated by the fact that this gain is the same obtained by Ungerboeck's one-dimensional TCM (1D-TCM), we search and tabulate optimum convolutional codes for BICM-T. We show that a properly design BICM system without interleaving performs asymptotically as well as 1D-TCM, and therefore, BICM-T should be considered as a good alternative for CM in nonfading channels. The main contribution of this paper is to present an analytical model for BICM-T which is used to explain the results presented in [8] and also to design a BICM-T system in nonfading channels.

II. SYSTEM MODEL AND PRELIMINARIES

Throughout this paper, we use boldface letters $\mathbf{c}_t = [c_{1,t}, \dots, c_{L,t}]$ to denote length- L row vectors and capital boldface letters $\mathbf{C} = [\mathbf{c}_1^T, \dots, \mathbf{c}_N^T]$ to denote matrices, where $(\cdot)^T$ denotes transposition. We use $d_H(\mathbf{C})$ to denote the total Hamming weight of the matrix \mathbf{C} . We denote

probability by $\Pr(\cdot)$ and the probability density function (pdf) of a random variable Λ by $p_\Lambda(\lambda)$. The convolution between two pdfs is denoted by $p_{\Lambda_1}(\lambda) * p_{\Lambda_2}(\lambda)$ and $\{p_\Lambda(\lambda)\}^{*w}$ denotes the w -fold self-convolution of the pdf $p_\Lambda(\lambda)$. A Gaussian distribution with mean value μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$, the Gaussian function with the same parameters by $\psi(\lambda; \mu, \sigma) \triangleq \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(\lambda-\mu)^2}{2\sigma^2})$, and the Q-function by $Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{u^2}{2}) du$. All the polynomial generators of the convolutional codes (CC) are given in octal notation.

A. System Model

The BICM system model under consideration is presented in Fig. 1. We use a constraint length K rate $R = \frac{1}{2}$ convolutional encoder connected to a 16-ary quadrature amplitude modulation (16-QAM) labeled by the binary reflected Gray code (BRGC) [9]. This configuration is indeed very simple yet practical yielding a spectral efficiency of two bits per real channel use. This example is not restrictive, of course, yet simplifies the presentation of the main ideas. The generalization to other modulations and coding rate is naturally possible but would obviously increase the complexity of notation potentially hindering the main concepts of the analysis presented in this paper.

The input sequence $\mathbf{i} = [i_1, \dots, i_N]$ is fed to the encoder (ENC) which at each time instant $t = 1, \dots, N$ generates two coded bits $\mathbf{c}_t = [c_{1,t}, c_{2,t}]$. We use the matrix $\mathbf{C} = [\mathbf{c}_1^T, \dots, \mathbf{c}_N^T]$ of size $2 \times N$ to represent the transmitted codeword. These coded bits are interleaved by Π , where the different interleaving alternatives will be discussed in detail in Sec. II-B. The coded and interleaved bits are then mapped to a 16-QAM symbol, where the 16-QAM constellation is formed by the direct product of two 4-ary pulse amplitude modulation (4-PAM) constellations labeled by the BRGC. Therefore, we analyze the real part of the constellation only, i.e., one of the constituent 4-PAM constellations. The mapper is defined as $\Phi : \{[11], [10], [00], [01]\} \rightarrow \{-3\Delta, -\Delta, \Delta, 3\Delta\}$, where we define

$$\Delta \triangleq \frac{1}{\sqrt{5}} \quad (1)$$

so that the PAM constellation normalized to unit average symbol energy, i.e., $E_s = 1$.

A quick inspection of the BRGC for 4-PAM reveals that the BRGC offers unequal error protection (UEP) to the transmitted bits depending on their position. In particular, the bit at

the first position ($k = 1$) receives higher protection¹ than the bit at the second position $k = 2$. More details about this can be found in [7]. Moreover, for $k = 2$ a bit labeled by zero (inner constellation points) will receive a lower protection than a bit labeled by one transmitted in the same bit position (outer constellation points), and therefore, the binary-input soft-output (BISO) channel for $k = 2$ is nonsymmetric. To simplify the analysis, we “symmetrize” the channel by randomly inverting the bits before mapping them to the 4-PAM symbol, i.e., $\tilde{\mathbf{C}} = \Pi(\mathbf{C}) \oplus \mathbf{S}$, where \oplus represents modulo-2 element-wise addition and the elements of the matrix $\mathbf{S} = [\mathbf{s}_1^T, \dots, \mathbf{s}_N^T] \in \{0, 1\}^{2 \times N}$ where $\mathbf{s}_t = [s_{1,t}, s_{2,t}]$ are randomly generated vectors of bits. Such a scrambling symmetrizes the BISO channel but it does not eliminate the UEP. We note that the scrambling is introduced only to simplify the analysis, and therefore, it is not shown in Fig. 1 nor used in the simulations. This symmetrization was in fact proposed in [5], and as we will see in Sec. IV, the bounds developed based on this symmetrization perfectly match the numerical simulations.

At each time $t = 1, \dots, N$, the coded and scrambled bits \tilde{c}_t are mapped to a symbol x_t , where $x_t = \Phi(\tilde{c}_t) \in \mathcal{X}$ and \mathcal{X} is the 4-PAM constellation. The symbols x_t are sent over an additive white Gaussian noise (AWGN) channel so the received signal is given by $y_t = x_t + z_t$, where z_t is a zero-mean Gaussian noise with variance $N_0/2$. The signal-to-noise ratio is defined as $\gamma \triangleq E_s/N_0 = 1/N_0$. At the receiver’s side, reliability metrics for the bits are calculated by the demapper Φ^{-1} in the form of logarithmic-likelihood ratios (L-values) as

$$\tilde{l}_{k,t} = \log \frac{\Pr(\tilde{c}_{k,t} = 1|y_t)}{\Pr(\tilde{c}_{k,t} = 0|y_t)}. \quad (2)$$

Since $\tilde{c}_{k,t} = c_{k,t} \oplus s_{k,t}$, it can be shown that $l_{k,t} = (-1)^{s_{k,t}} \tilde{l}_{k,t}$, i.e., after “descrambling”, the sign of the L-values is changed using $(-1)^{s_{k,t}}$. These L-values are deinterleaved and then passed to the decoder which calculates an estimate of the information sequence $\hat{\mathbf{i}}$.

B. The interleaver

Throughout this paper, three interleaving alternatives will be analyzed, cf. the block Π in Fig. 1. The first interleaving alternative is BICM with a single interleaver (BICM-S) introduced in [5].

¹The “protection” may be defined in different ways, where probably the simplest one is the bit error probability per bit position at the demapper’s output.

It is the most commonly used in the literature and corresponds to an interleaver that randomly permutes the bits \mathbf{C} prior to modulation, where the permutation is random in two “dimensions,” i.e., it permutes the bits over the bit positions and over time. The second alternative is BICM with multiple interleavers (M-interleavers, BICM-M) where the interleaver permutes the bits randomly only over time (and not over the bit positions). This can be seen as a particularization of the interleaver of BICM-S following an additional constraint: bits from the k th encoder’s output must be assigned to the k th modulator’s input. BICM-M was formally analyzed in [7] and in fact corresponds to the original model introduced by Zehavi in [4] (BICM) and Li in [10] (BICM with iterative decoding, BICM-ID). Recently, M-interleavers have also been proven to be asymptotically optimum for BICM-ID [11]. The last interleaving alternative, on which this paper focuses, is BICM with a trivial interleaver (BICM-T), i.e., when the interleaver Π in Fig. 1 is simply not present [8].

When BICM-T is considered, the resulting system is the one shown in Fig. 2. A careful examination of Fig. 2 reveals that the structure of the transmitter of BICM-T is the same as the transmitter of Ungerboeck’s one-dimensional TCM [2] or the TCM transmitter in [12, Fig. 4.17]. The transmitter of BICM-T can also be considered a particular case of the so-called “general TCM” [13, Fig. 18.11] when $k = \tilde{k}$ (using the notation of [13]) and when the BRGC is used instead of Ungerboeck’s set-partitioning. The receiver of BICM-T in Fig. 2 corresponds to a conventional BICM receiver, where L-values for each bit are computed and fed to a soft-input Viterbi decoder (VD). The difference between this receiver’s structure and a TCM receiver is that bit-level processing is used instead of a symbol-by-symbol VD. In conclusion, the BICM-T system introduced in [8] is simply a BRGC-based TCM transmitter used in conjunction with a BICM receiver. Nevertheless, through this paper, we use the name BICM-T to reflect the fact that this transmitter/receiver structure can be considered as a particular case of the BICM system in [4], [5], where the interleaver takes a trivial form.

C. The Decoder

A maximum likelihood sequence decoder (e.g., the VD) chooses the most likely coded sequence $\hat{\mathbf{C}}$ using the vector of channel observations $\mathbf{y} = [y_1, \dots, y_N]$ as

$$\hat{\mathbf{C}} = \max_{\mathbf{C} \in \mathcal{D}} \{ \log(\Pr\{\mathbf{C}|\mathbf{y}\}) \} \quad (3)$$

$$= \max_{\mathbf{C} \in \mathcal{D}} \left\{ \log \left(\prod_{t=1}^N \Pr\{\mathbf{c}_t|y_t\} \right) \right\}, \quad (4)$$

where \mathcal{D} is the set of all codewords, where to pass from (3) to (4) we used the memoryless property of the channel. If we assume that the bits $[c_{1,t}, c_{2,t}]$ are independent, we obtain

$$\log \left(\prod_{t=1}^N \Pr\{\mathbf{c}_t|y_t\} \right) = \log \left(\prod_{k=1}^2 \prod_{t=1}^N \Pr\{c_{k,t}|\mathbf{y}\} \right). \quad (5)$$

Under this independence assumption and by using the relation between an L-value l and the bit's probabilities of being $b \in \{0, 1\}$

$$\Pr\{b|y\} = \frac{e^{bl}}{1 + e^l}, \quad (6)$$

we obtain

$$\begin{aligned} \log \left(\prod_{k=1}^2 \prod_{t=1}^N \Pr\{c_{k,t}|\mathbf{y}\} \right) &= \sum_{k=1}^2 \sum_{t=1}^N \log(\Pr\{c_{k,t}|\mathbf{y}\}) \\ &= \sum_{k=1}^2 \sum_{t=1}^N c_{k,t} l_{k,t} - \sum_{k=1}^2 \sum_{t=1}^N \log(1 + \exp(l_{k,t})). \end{aligned} \quad (7)$$

Since the second term in (7) is independent of \mathbf{C} , it is irrelevant to the decision of the decoder in (4). Therefore, the final decision of the decoder can be written as

$$\hat{\mathbf{C}} = \max_{\mathbf{C} \in \mathcal{D}} \left\{ \sum_{k=1}^2 \sum_{t=1}^N c_{k,t} l_{k,t} \right\}. \quad (8)$$

In a BICM system with convolutional codes, the decoder is implemented using an off-the-shelf soft-input VD, which assumes that the bits are independent, and thus, uses the relation in (5) (i.e., it uses the decision rule in (8)). The relation in (5) is in indeed valid when BICM-S [5] or BICM-M [4], [7], [11] configurations are used, since in those cases, the use of a random interleaver (cf. Sec. II-B) assure that the bits $[c_{1,t}, c_{2,t}]$ are transmitted in different symbols, and therefore, are affected by different noise realizations.

However, when BICM-T with a soft-input VD is considered, and since the bits $[c_{1,t}, c_{2,t}]$ are affected by the same noise realization, the relation in (5) does not hold, i.e., the two L-values

passed to the decoder at any time instant t are not independent. Nevertheless, the decoder treats the bits as independent and still uses the decision rule in (8). In principle, it would be possible to design a decoder for BICM-T that takes into account this inconsistency, i.e., a decoder that does not assume independent bits. However, this is out of the scope of this paper and would also go against the flexibility offered by BICM. Moreover, we will show in the following section that even with this inconsistency, BICM-T in nonfading channels outperforms BICM-S and BICM-M.

III. PERFORMANCE EVALUATION

A. BER Performance

Because of the symmetrization of the channel, we can, without loss of generality, assume that the all-zero codeword was transmitted. We define \mathcal{E} as the set of codewords corresponding to paths in the trellis of the code diverging from the zero-state at the arbitrarily chosen instant $t = t_0$, and remerging with it after T trellis stages. We also denote these codewords as $\mathbf{E} \triangleq [e_1^T, \dots, e_T^T]$, where $e_t = [e_{1,t}, e_{2,t}]$. Then, the bit error rate (BER) can be upper-bounded using a union bound (UB) as

$$\text{BER} \leq \text{UB} \triangleq \sum_{\mathbf{E} \in \mathcal{E}} \text{PEP}(\mathbf{E}) d_H(\mathbf{i}_{\mathbf{E}}), \quad (9)$$

where $d_H(\mathbf{i}_{\mathbf{E}})$ is the Hamming weight of the input sequence $\mathbf{i}_{\mathbf{E}}$ corresponding to the codeword \mathbf{E} , and the pairwise error probability (PEP) is given by (cf. (8))

$$\text{PEP}(\mathbf{E}) = \Pr \left\{ \sum_{t=t_0}^{t_0+T-1} (e_{1,t} l_{1,t} + e_{2,t} l_{2,t}) > 0 \right\}. \quad (10)$$

The general expression for the PEP in (10) and the UB in (9) reduce to well-known particular cases if simplifying assumptions for the distribution of $l_{k,t}$ are adopted.

1) Independent and identically distributed L-values (BICM-S): In BICM-S, the L-values $l_{k,t}$ passed to the decoder are independent and identically distributed (i.i.d.). They can be described using the conditional pdf $p(\lambda|b)$ with $b \in \{0, 1\}$ and where the pdf is independent of k and t . In this case, the PEP in (10) depends only on the Hamming weight of the codeword \mathbf{E} , i.e.,

$$\begin{aligned} \text{PEP}(\mathbf{E}) &= \text{PEP}_S(d_H(\mathbf{E})) \\ &= \int_0^\infty \{p(\lambda|b=0)\}^{*d_H(\mathbf{E})} d\lambda. \end{aligned} \quad (11)$$

The UB in (9) can be expressed as

$$\text{UB}_S = \sum_w \text{PEP}_S(w) \sum_{\mathbf{E} \in \mathcal{D}_w} d_H(\mathbf{i}_{\mathbf{E}}) \quad (12)$$

$$= \sum_w \text{PEP}_S(w) \beta_w^C, \quad (13)$$

where \mathcal{D}_w represents the set of codewords with Hamming weight w , i.e., $\mathcal{D}_w \triangleq \{\mathbf{E} \in \mathcal{E} : d_H(\mathbf{E}) = w\}$. To pass from (12) to (13) we group the codewords \mathbf{E} that have the same Hamming weight and add their contributions, which results in the well-known weight distribution spectrum of the code β_w^C . The expression in (13) is the most common expression for the UB for BICM, cf. [5, eq. (26)], [6, eq. (4.12)].

2) *Independent but not identically distributed L-values (BICM-M)*: In BICM-M, the L-values passed to the each decoder's input are independent, however, their conditional pdf depends on the bit's position $k = 1, 2$. Thus, the L-values are modeled by the set of conditional pdfs $\{p_1(\lambda|b), p_2(\lambda|b)\}$. The PEP in this case is given by

$$\begin{aligned} \text{PEP}(\mathbf{E}) &= \text{PEP}_M(\bar{w}_{\mathbf{E},1}, \bar{w}_{\mathbf{E},2}) \\ &= \int_0^\infty \{p_1(\lambda|b_1 = 0)\}^{*\bar{w}_{\mathbf{E},1}} * \{p_2(\lambda|b_2 = 0)\}^{*\bar{w}_{\mathbf{E},2}} d\lambda, \end{aligned} \quad (14)$$

where $\bar{w}_{\mathbf{E},k}$ is the Hamming weight of the k th row of \mathbf{E} . The UB in (9) can be expressed as

$$\begin{aligned} \text{UB}_M &= \sum_{w_1, w_2} \text{PEP}_M(w_1, w_2) \sum_{\mathbf{E} \in \mathcal{D}_{w_1, w_2}} d_H(\mathbf{i}_{\mathbf{E}}) \\ &= \sum_{w_1, w_2} \text{PEP}_M(w_1, w_2) \beta_{w_1, w_2}^C, \end{aligned} \quad (15)$$

where \mathcal{D}_{w_1, w_2} is the set of codewords with *generalized* Hamming weight $[w_1, w_2]$ (w_k in its k th row), i.e., $\mathcal{D}_{w_1, w_2} \triangleq \{\mathbf{E} \in \mathcal{E} : w_1 = \bar{w}_{\mathbf{E},1}, w_2 = \bar{w}_{\mathbf{E},2}\}$, and β_{w_1, w_2}^C is the generalized weight distribution spectrum of the code that takes into account the errors at each encoder's output separately. The UB in (15) was shown in [7] to be useful when analyzing the UEP introduced by the binary labeling and also to optimize the interleaver and the code.

3) *BICM without bit-interleaving (BICM-T)*: For BICM-T, yet a different particularization of (10) must be adopted. Let $\Lambda_{\mathbf{E}}$ be the metric associated to the codeword \mathbf{E} and assume without loss of generality that $t_0 = t$. This metric is a sum of independent random variables, i.e.,

$$\Lambda_{\mathbf{E}} \triangleq \Lambda_t + \Lambda_{t+1} + \Lambda_{t+2} + \dots, \quad (16)$$

where $\Lambda_t = e_{1,t}l_{1,t} + e_{2,t}l_{2,t}$ corresponds to the elements defining the PEP in (10). We then express the t th metric as

$$\Lambda_t(\mathbf{e}_t, \mathbf{s}_t) = \begin{cases} 0, & \text{if } \mathbf{e}_t = [0, 0] \\ (-1)^{s_{1,t}} \tilde{l}_{1,t}, & \text{if } \mathbf{e}_t = [1, 0] \\ (-1)^{s_{2,t}} \tilde{l}_{2,t}, & \text{if } \mathbf{e}_t = [0, 1] \\ \sum_{k=1}^2 (-1)^{s_{k,t}} \tilde{l}_{k,t}, & \text{if } \mathbf{e}_t = [1, 1] \end{cases}, \quad (17)$$

where we use $\Lambda_t(\mathbf{e}_t, \mathbf{s}_t)$ to show that Λ_t depends on the scrambling's outcome \mathbf{s}_t (through $\tilde{l}_{k,t}$) and the error pattern at time t , \mathbf{e}_t .

Since $\tilde{l}_{k,t}$ are random variables (that depend on k and x_t), according to (17), there exist three pdfs that can be used to model the individual metrics in (16). We denote the set of these three conditional pdfs by $\{p_1(\lambda|b_1), p_2(\lambda|b_2), p_\Sigma(\lambda|\mathbf{b})\}$, for the three relevant cases defined in (17), respectively. We note that $p_\Sigma(\lambda|\mathbf{b})$ is conditioned not only on one bit, but on the pair of transmitted bits $\mathbf{b} = [b_1, b_2]$, where b_1, b_2 , and \mathbf{b} represent the bits $c_{1,t}, c_{2,t}$, and \mathbf{c}_t , respectively. From (16), and due to the independence of the individual metrics, the PEP in (10) can be expressed as

$$\begin{aligned} \text{PEP}(\mathbf{E}) &= \text{PEP}_T(w_{\mathbf{E},1}, w_{\mathbf{E},2}, w_{\mathbf{E},\Sigma}) \\ &= \int_0^\infty \{p_1(\lambda|b_1 = 0)\}^{*w_{\mathbf{E},1}} * \{p_2(\lambda|b_2 = 0)\}^{*w_{\mathbf{E},2}} * \{p_\Sigma(\lambda|\mathbf{b} = [0, 0])\}^{*w_{\mathbf{E},\Sigma}} d\lambda, \end{aligned} \quad (18)$$

where $w_{\mathbf{E},k}$ is the number of columns in \mathbf{E} where only the k th row of \mathbf{E} is one, and $w_{\mathbf{E},\Sigma}$ is the number columns in \mathbf{E} where both entries are equal to one. Clearly

$$d_H(\mathbf{E}) = w_{\mathbf{E},1} + w_{\mathbf{E},2} + 2w_{\mathbf{E},\Sigma}. \quad (19)$$

Example 1 (Error event at minimum Hamming distance of (5, 7) code): Consider the constraint length $K = 3$ optimum distance spectrum convolutional code (ODSCC) with polynomial generators (5, 7) [14, Table I]. The free distance of the code is $d_H^{\text{free}} = 5$, and $\beta_5^C = 1$, i.e., there is one divergent path at Hamming distance five from the all-zero codeword, and the Hamming weight of that path is $d_H(\mathbf{i}_{\mathbf{E}}) = 1$. Moreover, it is possible to show that this codeword is

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

i.e., $d_H(\mathbf{E}) = 5$, $w_{\mathbf{E},1} = 0$, $w_{\mathbf{E},2} = 1$, and $w_{\mathbf{E},\Sigma} = 2$. Also, $\bar{w}_{\mathbf{E},1} = 2$ and $\bar{w}_{\mathbf{E},2} = 3$.

We define $\mathcal{D}_{w_1, w_2, w_\Sigma}$ as the set of codewords \mathbf{E} with w_1 columns such that $\mathbf{e}_t = [1, 0]$, w_2 columns with $\mathbf{e}_t = [0, 1]$, and w_Σ columns with $\mathbf{e}_t = [1, 1]$, i.e., $\mathcal{D}_{w_1, w_2, w_\Sigma} \triangleq \{\mathbf{E} \in \mathcal{E} : w_1 = w_{\mathbf{E},1}, w_2 = w_{\mathbf{E},2}, w_\Sigma = w_{\mathbf{E},\Sigma}\}$. Using this, the UB expression in (9) for BICM-T is given by

$$\begin{aligned} \text{UB}_T &= \sum_{w_1, w_2, w_\Sigma} \text{PEP}_T(w_1, w_2, w_\Sigma) \sum_{\mathbf{E} \in \mathcal{D}_{w_1, w_2, w_\Sigma}} d_H(\mathbf{i}_{\mathbf{E}}) \\ &= \sum_{w_1, w_2, w_\Sigma} \text{PEP}_T(w_1, w_2, w_\Sigma) \beta_{w_1, w_2, w_\Sigma}^{\mathcal{C}}, \end{aligned} \quad (20)$$

where $\beta_{w_1, w_2, w_\Sigma}^{\mathcal{C}}$ is a weight distribution spectrum of the code \mathcal{C} that not only considers the generalized weight $[w_1, w_2]$ of the codewords, but takes into account the temporal behavior, i.e., it considers the case when $\mathbf{e}_t = [1, 1]$ as a different kind of event. This differs from $\beta_{w_1, w_2}^{\mathcal{C}}$, where such an event will be simply considered as an extra contribution to the total generalized weight.

B. PDF of the L-values

In order to calculate the PEP for BICM-T in (18) we need to compute the set of conditional pdfs $\{p_1(\lambda|b_1), p_2(\lambda|b_2), p_\Sigma(\lambda|\mathbf{b})\}$. In this subsection we show how to find approximations for these PDFs.

The L-values in (2) can be expressed as

$$\tilde{l}_{k,t} = \log \frac{\sum_{x \in \mathcal{X}_{k,1}} p(y_t|x)}{\sum_{x \in \mathcal{X}_{k,0}} p(y_t|x)}, \quad (21)$$

where $\mathcal{X}_{k,b}$ is the set of constellation symbols labeled with b at bit position k . Using the fact that the channel is Gaussian and if the so-called max-log approximation $\log(e^a + e^b) \approx \max\{a, b\}$ is used, the L-values can be expressed as

$$\tilde{l}_{k,t}(y_t|\mathbf{s}_t) \approx \gamma \left[\min_{x \in \mathcal{X}_{k,0}} (y_t - x)^2 - \min_{x \in \mathcal{X}_{k,1}} (y_t - x)^2 \right], \quad (22)$$

where from now on we use the notation $\tilde{l}_{k,t}(y_t|\mathbf{s}_t)$ to emphasize that the L-values depend on the received signal and the scrambler's outcome \mathbf{s}_t . In fact, the L-values depend on the transmitted symbol x_t , however, and since $\mathbf{c}_t = \mathbf{0}$ and no interleaving is performed, x_t is completely determined by \mathbf{s}_t .

The L-value in (22) is a piece-wise linear function of y_t . Moreover, the L-values $\Lambda_t(e_t, \mathbf{s}_t)$ in (17) are linear combinations of $\tilde{l}_{k,t}(y_t|\mathbf{s}_t)$ in (22), and therefore, they are also piece-wise linear

functions of y_t . Two cases are of particular interest, namely, when $\mathbf{e}_t = [1, 0]$ or $\mathbf{e}_t = [0, 1]$, and when $\mathbf{e}_t = [1, 1]$. The piece-wise linear relationships for first case are shown in in Fig. 3 a) for 4-PAM. In this figure we also show the constellation symbols and we use the notation $\mathbf{s}_t = [0/1, :]$ and $\mathbf{s}_t =[:, 0/1]$ to show that for $\mathbf{e}_t = [1, 0]$ and $\mathbf{e}_t = [0, 1]$ the L-values $\Lambda_t(\mathbf{e}_t, \mathbf{s}_t)$ are independent of $s_{2,t}$ and $s_{1,t}$, respectively. In Fig. 3 b), the four possible cases when $\mathbf{e}_t = [1, 1]$ are shown.

For a given transmitted symbol x_t (determined by \mathbf{s}_t), the received signal y_t is a Gaussian random variable with mean x_t and variance $N_0/2$. Therefore, each L-value $\Lambda_t(\mathbf{e}_t, \mathbf{s}_t)$ in (17) is a sum of piece-wise Gaussian functions². In order to obtain expressions that are easy to work with, we use the so-called zero-crossing approximation of the L-values proposed in [15, Sec. III-C] which replaces all the Gaussian pieces required in the max-log model of L-values by a single Gaussian function. Intuitively, this approximation states that

$$\Lambda_t(y_t|\mathbf{e}_t, \mathbf{s}_t) \approx \hat{a}(\mathbf{e}_t, \mathbf{s}_t)y_t + \hat{b}(\mathbf{e}_t, \mathbf{s}_t), \quad (23)$$

where $\hat{a}(\mathbf{e}_t, \mathbf{s}_t)$ and $\hat{b}(\mathbf{e}_t, \mathbf{s}_t)$ are the slope and the free coefficient of the closest linear piece to the transmitted symbol x_t .

In Table I we show the values of $\hat{a}(\mathbf{e}_t, \mathbf{s}_t)$ and $\hat{b}(\mathbf{e}_t, \mathbf{s}_t)$ defining (23) for 4-PAM, where for notation simplicity we have defined

$$\alpha \triangleq 4\gamma\Delta^2. \quad (24)$$

To clarify how these coefficients are obtained, consider for example $\mathbf{e}_t = [0, 1]$. In this case, for $\mathbf{s}_t = [1, 1]$, which corresponds to $x_t = -3\Delta$, the closest linear piece intersecting the x -axis is the left-most part of the curve labeled in Fig. 3 by $\mathbf{e}_t = [0, 1]$ and $\mathbf{s}_t =[:, 1]$ (dashed-dotted line). If for example $\mathbf{e}_t = [0, 1]$ and $\mathbf{s}_t = [0, 0]$ ($x_t = \Delta$), the closest linear piece is the right-most piece labeled by $\mathbf{e}_t = [0, 1]$ and $\mathbf{s}_t =[:, 0]$ (dashed line). All the other values in Table I can be found by a similar direct inspection of Fig. 3.

Using the approximation in (23), the L-values can be modeled as Gaussian random variables where their mean and variance depend on \mathbf{s}_t , γ , and \mathbf{e}_t , i.e.,

$$p_{\Lambda_t}(\lambda|\mathbf{e}_t, \mathbf{s}_t) = \psi(\lambda; \hat{\mu}(\mathbf{e}_t, \mathbf{s}_t), \hat{\sigma}^2(\mathbf{e}_t, \mathbf{s}_t)), \quad (25)$$

²Closed-form expressions for these pdfs of $\Lambda_t(\mathbf{e}_t, \mathbf{s}_t)$ when $\mathbf{e}_t = [1, 0]$ and $\mathbf{e}_t = [1, 1]$ (cf. Fig. 3 a) were presented in [15].

where the mean value and variance are given by

$$\hat{\mu}(\mathbf{e}_t, \mathbf{s}_t) = x_t \hat{a}(\mathbf{e}_t, \mathbf{s}_t) + \hat{b}(\mathbf{e}_t, \mathbf{s}_t) \quad (26)$$

$$\hat{\sigma}^2(\mathbf{e}_t, \mathbf{s}_t) = [\hat{a}(\mathbf{e}_t, \mathbf{s}_t)]^2 \frac{N_0}{2}. \quad (27)$$

In Table II we show the obtained mean values and variances for the same cases presented in Table I.

To obtain the pdf of Λ_t in (17), we simply average (25) over the symbols, which are assumed to be equiprobable. This results in the following expression

$$p_{\Lambda_t}(\lambda) = \begin{cases} \frac{1}{2} [\psi(\lambda; -3\alpha, 2\alpha) + \psi(\lambda; -\alpha, 2\alpha)], & \text{if } \mathbf{e}_t = [1, 0] \\ \psi(\lambda; -\alpha, 2\alpha), & \text{if } \mathbf{e}_t = [0, 1] \\ \psi(\lambda; -4\alpha, 8\alpha), & \text{if } \mathbf{e}_t = [1, 1] \end{cases} \quad (28)$$

IV. DISCUSSION AND APPLICATIONS

In the previous section, we developed approximations for the pdf of the L-values passed to the decoder in BICM-T. In this section we use them to quantify the gains offered by BICM-T over BICM-S, to define asymptotically optimum CCs, and to compare BICM-T with Ungerboeck's 1D-TCM.

A. Performance of BICM-T

Expression (28) show the pdf of the L-values needed to compute the UB of BICM-T, cf. (18) and (20). Moreover, due to the simplifications introduced in the previous subsections the results in (28) only involve Gaussian pdfs, which greatly simplifies the PEP computation in (18).

Theorem 1: The UB for BICM-T is

$$\text{UB}_T = \sum_{w_1, w_2, w_\Sigma} \beta_{w_1, w_2, w_\Sigma}^c \left(\frac{1}{2}\right)^{w_1} \sum_{j=0}^{w_1} \binom{w_1}{j} Q \left(\sqrt{\frac{(w_1 + w_2 + 4w_\Sigma + 2j)^2 2\gamma}{(w_1 + w_2 + 4w_\Sigma) 5}} \right). \quad (29)$$

Proof: Inserting (28) in (18), changing the convolution of sums into a sum of convolutions, and using $\psi(\lambda; \mu_1, \sigma_1^2) * \dots * \psi(\lambda; \mu_J, \sigma_J^2) = \psi(\lambda; \sum_{j=1}^J \mu_j, \sum_{j=1}^J \sigma_j^2)$, the PEP in (18) can be expressed as

$$\text{PEP}_T(w_1, w_2, w_\Sigma) = \int_0^\infty \left(\frac{1}{2}\right)^{w_1} \sum_{j=0}^{w_1} \binom{w_1}{j} \psi(\lambda; \mu_{1,2,\Sigma,j}, \sigma_{1,2,\Sigma}^2) d\lambda, \quad (30)$$

where

$$\mu_{1,2,\Sigma,j} = -(w_1 + w_2 + 4w_\Sigma + 2j)\alpha \quad (31)$$

$$\sigma_{1,2,\Sigma}^2 = 2(w_1 + w_2 + 4w_\Sigma)\alpha. \quad (32)$$

By using the definition of α in (24) and Δ in (1), and (31) and (32) in (30), and the UB definition in (20), the expression in (29) is obtained. ■

In Fig. 4, numerical results for BICM-T with 4-PAM labeled with the BRGC and using the ODSCCs (5, 7) ($K = 3$) and (247, 371) ($K = 8$) [14, Table I] are shown. For BICM-M two configurations are considered for each code. The first one is when all the bits from the first encoder's output are assigned to the first modulator's input and all the bits from the second encoder's output are sent to the second modulator's input. The second alternative simply corresponds to the opposite, i.e., all the bits from the first encoder's output are sent over $k = 2$ and the bits from the second encoder's output are sent over $k = 1$. This is equivalent to defining the code by simply swapping the order of the polynomial generators. For these two particular codes, the configuration that minimizes the BER for medium to high SNR is the second one, i.e., when all the bits generated by the polynomial (7) or (371) are sent over $k = 1$ and all the bits generated by the polynomial (5) or (247) are sent over $k = 2$. We denote the configuration that minimizes (or maximizes) the BER by “Best” (or “Worst”).

To compute the UB for BICM-S and BICM-M, we use the expressions in [7, eq. (22)–(23)], and for BICM-T we use Theorem 1. All the UB computations were carried out considering a truncated spectrum of the code, i.e., $\{w, w_1, w_2, w_\Sigma\} \leq 30$ which is calculated numerically using a breadth first search algorithm [16]. The results in Fig. 4 show that the UB developed in this paper for BICM-T predict well the simulation results. Also, these results show that for these particular codes, the gains obtained by using BICM-M instead of BICM-S are small, although larger gains were obtained in [7] for other codes/configurations. On the other hand, the gains by using BICM-T instead of BICM-S for a BER target of 10^{-7} are approximately 2 dB for $K = 3$ and 1 dB for $K = 8$. Moreover, these gains are obtained by decreasing the complexity of the system, i.e., by not doing interleaving/de-interleaving.

B. Asymptotic Performance

In this subsection, we analyze the performance of BICM-T for asymptotically high SNR and we compare it with BICM-S.

Theorem 2: The asymptotic UB for BICM-T and a given code \mathcal{C} can be expressed as

$$\text{UB}'_{\text{T}} = M^{\mathcal{C}} Q \left(\sqrt{\frac{2\gamma A^{\mathcal{C}}}{5}} \right), \quad (33)$$

where

$$A^{\mathcal{C}} \triangleq \min_{\substack{w_1, w_2, w_{\Sigma} \\ \beta_{w_1, w_2, w_{\Sigma}}^{\mathcal{C}} \neq 0}} (w_1 + w_2 + 4w_{\Sigma}) \quad (34)$$

$$M^{\mathcal{C}} = \sum_{\substack{w_1, w_2, w_{\Sigma} \\ \beta_{w_1, w_2, w_{\Sigma}}^{\mathcal{C}} \neq 0 \\ w_1 + w_2 + 4w_{\Sigma} = A}} \beta_{w_1, w_2, w_{\Sigma}}^{\mathcal{C}} \left(\frac{1}{2} \right)^{w_1}. \quad (35)$$

Proof: The UB in (29) is a sum of weighted Q-functions. For high SNR, and for each (w_1, w_2, w_{Σ}) there is a Q-function that dominates the inner sum in (29). This is obtained for $j = 0$, which completes the proof. ■

For comparison purposes, we present here the performance of BICM-S at asymptotically high SNR. This can be obtained for example by particularizing [7, eq. (25)] to the conventional BICM configuration with one single interleaver. The asymptotic performance of BICM-S is given by

$$\text{UB}'_{\text{S}} = \left(\frac{3}{4} \right)^{d_{\text{H}}^{\text{free}}} \beta_{d_{\text{H}}^{\text{free}}}^{\mathcal{C}} Q \left(\sqrt{\frac{2d_{\text{H}}^{\text{free}}\gamma}{5}} \right), \quad (36)$$

where $d_{\text{H}}^{\text{free}}$ is the free Hamming distance of the code which can be expressed as $d_{\text{H}}^{\text{free}} = w_1^{\text{free}} + w_2^{\text{free}} + 2w_{\Sigma}^{\text{free}}$, cf. (19).

In Fig. 4, we show asymptotic UBs for $K = 3$. For BICM-T we used Theorem 2, for BICM-S we use (36), and for BICM-M we use [7, eq. (25)]. All of them are shown to follow the simulation results quite well. Similar results can be obtained for the code with $K = 8$, however, we do not show those results not to overcrowd the figure.

The asymptotic gain (AG) obtained by using BICM-T instead of BICM-S is obtained directly from Theorem 2 and (36), as stated in the following corollary.

Corollary 3: The AG obtained by using BICM-T instead of BICM-S

$$\text{AG}_{\text{S} \rightarrow \text{T}} = 10 \log_{10} \left(\frac{A^{\mathcal{C}}}{w_1^{\text{free}} + w_2^{\text{free}} + 2w_{\Sigma}^{\text{free}}} \right). \quad (37)$$

Example 2 (AG for the (5, 7) code): For the particular code (5, 7), it is possible to see that the solution of (34) corresponds to the event at minimum Hamming distance³, i.e., $d_H^{\text{free}} = 5$, $w_{\mathbf{E},1} = 0$, $w_{\mathbf{E},2} = 1$, $w_{\mathbf{E},\Sigma} = 2$ (cf. Example 1), and therefore, $A^C = 9$. This result in an AG of $10 \log_{10} \left(\frac{9}{5} \right) \approx 2.55$ dB. Moreover, since the input sequence that generates the codeword at minimum Hamming distance has Hamming weight one ($\beta_{0,1,2}^C = 1$), we obtain $M_{0,1,2}^C = 1$ for the configuration “Worst”. If the polynomials are swapped (which corresponds to swapping the rows of \mathbf{E}), i.e., if we consider the code (7, 5), we obtain $w_{\mathbf{E},1} = 1$, $w_{\mathbf{E},2} = 0$, $w_{\mathbf{E},\Sigma} = 2$ and the same A^C (since A^C does not depend on the order of the polynomials). However, in this case $M_{1,0,2}^C = 1/2$. These two asymptotic bounds are shown in Fig. 4, where the influence of the coefficient M^C can be observed in both numerical results and asymptotic bounds.

C. Asymptotically Optimum Convolutional Codes

Optimum CCs are usually defined in terms of minimum distance, i.e., good CCs are the one that for a given rate and constraint length have the maximum free distance (MFD) [17, Sec. 8.2.5]. The MFD criterion can be refined if the multiplicities associated to the different weights are considered [14], [13, Sec. 12.3]. This optimality criterion resulted in the ODSCCs which are optimal in both binary transmission and in BICM-S, cf. (36). for BICM-M, we have shown in [7] that d_H^{free} is still a good indicator of the optimality of the code (as well as its multiplicity), however, a generalized weight distribution spectrum of the code should be considered, cf. β_{w_1, w_2}^C in (15). If BICM-T is considered, and as a direct consequence of Theorem 2, asymptotically optimum convolutional codes (AOCCs) can be defined.

Definition 1 (Asymptotically optimum convolutional codes for BICM-T): A CC is said to be an AOCC if among all codes with the same K and $R = 1/2$ it has the highest A^C , and among all codes with the same K and $R = 1/2$ it has the lowest multiplicity M^C .

We have performed an exhaustive numerical search for AOCCs based on Definition 1. We considered for constraint lengths $K = 3, 4, \dots, 8$ and all codes with free distance $0 < d_H^{\text{free}} \leq \hat{d}_H^{\text{free}}$, where \hat{d}_H^{free} is the free distance of the ODSCC. The spectrum was truncated as $w_1 + w_2 + 4w_\Sigma \leq \hat{d}_H^{\text{free}} + 8$ and the search was performed in lexicographic order. The results are shown in Table III, where we also include the ODSCCs for comparison. If there exist more than one AOCC for a

³However, this is not always true for other codes.

given K , we present the first one in the list. These results show that in general the minimum free distance of the code is not the proper criterion in BICM-T, i.e., codes that are not MFD codes perform better than the ODSCCs, cf. $K = 6, 7$. In fact, only for $K = 3$ the ODSCC is also optimum for BICM-T⁴. In this table we also present the AG that BICM-T offers with respect to BICM-S. The values obtained are around 2 dB.

In Fig. 5, we show the values of A^C and M^C for all the possible codes with $K = 5$. This figure shows that for $K = 5$ the AOCC and the ODSCCs have the same asymptotic performance (same A^C), however, the multiplicity of the AOCC is smaller. In Fig. 6, we show similar results for $K = 6$, where we only show a subset of all the possible codes. This figure shows that for $K = 6$, the ODSCC gives a worse (smaller A^C) asymptotic performance compared to the AOCC. Moreover, the AOCC code in this case has $d_H^{\text{free}} = 7$ while the ODSCC has $d_H^{\text{free}} = 8$, cf. Table III. The same phenomenon occurs for $K = 7$.

D. BICM-T vs. TCM

As mentioned in Sec. II-B, the transmitter of BICM-T is identical to the transmitter of Ungerboeck's 1D-TCM. In this subsection we compare their asymptotic performance.

We have previously defined in (37) the AG of BICM-T over BICM-S. It is also possible to define the AG of BICM-T compared to uncoded transmission with the same spectral efficiency (uncoded 2-PAM). Since the minimum squared Euclidean distance of the 2-PAM constellation is 4, the AG is given by

$$\text{AG}_{\text{UC} \rightarrow \text{T}} = 10 \log_{10} \left(\frac{A^C}{5} \right). \quad (38)$$

The AG in (38) is tabulated in the last column of Table III. For $K = 3$, $\text{AG}_{\text{UC} \rightarrow \text{T}}$ is equal to 2.55 dB, which is the same as $\text{AG}_{\text{S} \rightarrow \text{T}}$. This is because BICM-S with $K = 3$ does not offer any AG compared to uncoded 2-PAM. Analyzing the results in the last column of Table III, we find that they are the same ones obtained by 1D-TCM, cf. [2, Table I]. This simply states that if BICM-T is used with the correct CC, it performs asymptotically as well as 1D-TCM, and therefore, it should be considered as good alternative for CM in nonfading channels. However, this is not the case if BICM-S is used, or if BICM-T is used with the ODSCCs.

⁴For $K = 4$ the AOCC (13, 17) has, in fact, the same spectrum $\beta_{w_1, w_2, w_\Sigma}^C$ than the ODSCC (15, 17). The AOCC appears in the list because of the lexicographic order search.

V. CONCLUSIONS

In this paper, we gave a formal explanation of why gains can be obtained when BICM-T is used in nonfading channels. BICM-T was shown to be a TCM transmitter used with a BICM receiver. An analytical model was developed and a new type of distance spectrum for the code was introduced, which is the relevant characteristic to optimize CCs for BICM-T. The analytical model was used to validate the numerical results and to show that the use of the ODSCCs, which rely on the regular minimum free distance criterion, is suboptimal.

For simplicity, the analysis presented in this paper was done only for a simple BICM configuration, and therefore, it is still unknown what the performance gains will be in a more general setup, e.g., when the number of encoder's outputs is not the same as the modulator's input, for different spectral efficiencies, or when a less trivial (but still not infinitely long and random) interleaver is used. All these questions are left for further investigation.

REFERENCES

- [1] J. L. Massey, "Coding and modulation in digital communications," in *International Zürich Seminar on Digital Communications*, Zürich, Switzerland, Mar. 1974.
- [2] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.
- [3] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
- [4] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.
- [5] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [6] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.
- [7] A. Alvarado, E. Agrell, L. Szczecinski, and A. Svensson, "Exploiting UEP in QAM-based BICM: Interleaver and code design," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 500–510, Feb. 2010.
- [8] C. Stierstorfer, R. Fischer, and J. Huber, "Optimizing BICM with convolutional codes for transmission over the AWGN channel," in *2010 International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2010.
- [9] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.
- [10] X. Li and J. Ritcey, "Bit-interleaved coded modulation with iterative decoding using soft feedback," *Electronic Letters*, vol. 34, no. 10, pp. 942–943, May 1998.
- [11] A. Alvarado, L. Szczecinski, E. Agrell, and A. Svensson, "On BICM-ID with multiple interleavers," *IEEE Commun. Lett.*, vol. 14, no. 9, Sep. 2010.
- [12] S. H. Jamali and T. Le-Ngoc, *Coded-Modulation Techniques for Fading Channels*. Kluwer Academic Publishers, 1994.
- [13] S. Lin and D. J. Costello, Jr., *Error Control Coding*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, Inc., 2004.
- [14] P. Frenger, P. Orten, and T. Ottosson, "Convolutional codes with optimum distance spectrum," *IEEE Trans. Commun.*, vol. 3, no. 11, pp. 317–319, Nov. 1999.
- [15] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2 -QAM: Closed-form approximations and applications," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2071–2079, July 2009.
- [16] J. Belzile and D. Haccoun, "Bidirectional breadth-first algorithms for the decoding of convolutional codes," *IEEE Trans. Commun.*, vol. 41, no. 2, pp. 370–380, Feb. 1993.
- [17] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2000.

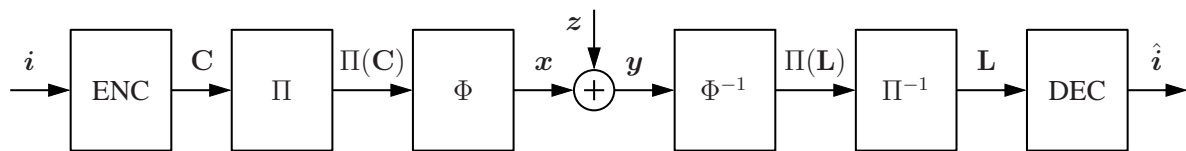


Fig. 1. Model of BICM transmission.

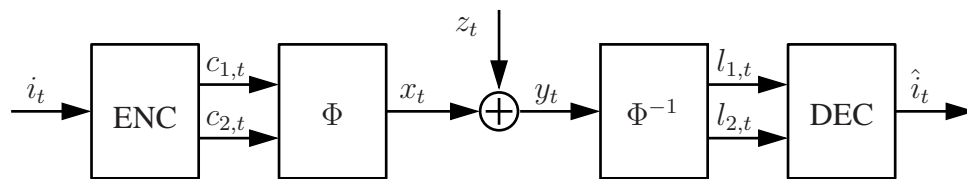


Fig. 2. BICM-T system analyzed in this paper for any time instant t .

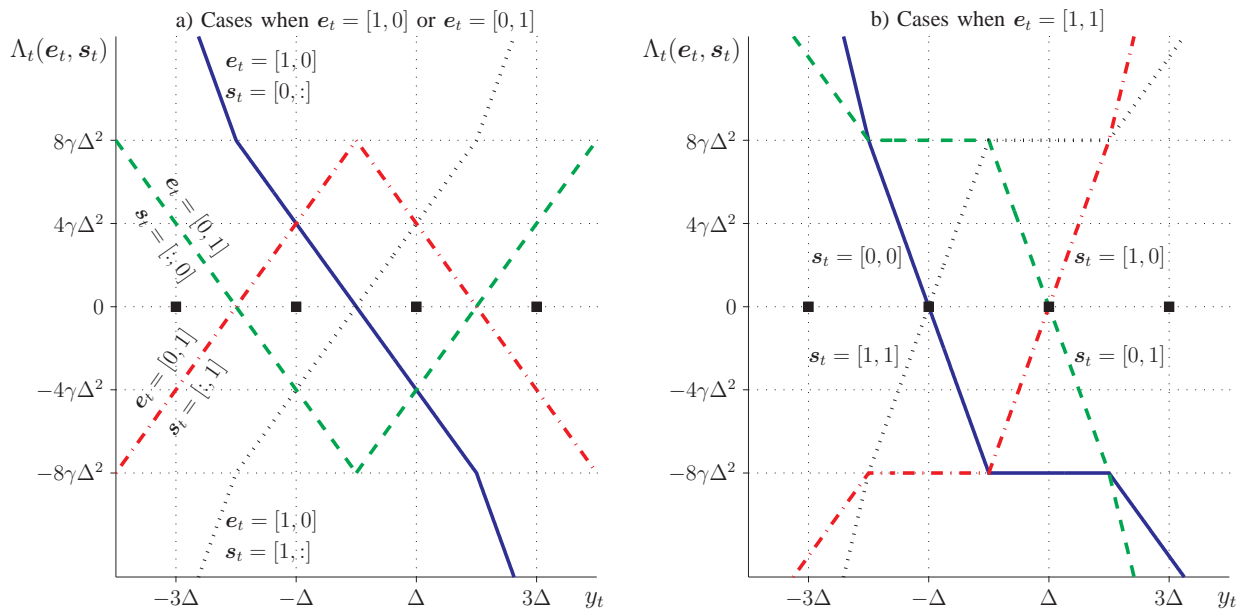


Fig. 3. Piece-wise relation between the L-values $\Lambda_t(e_t, s_t)$ in (17) and the received signal y_t for 4-PAM for all the possible values of e_t and s_t . The relation for the case when $e_t = [1, 0]$ or $e_t = [0, 1]$ is shown in a), and the relation when $e_t = [1, 1]$ is shown in b). The transmitted symbols are shown with black squares.

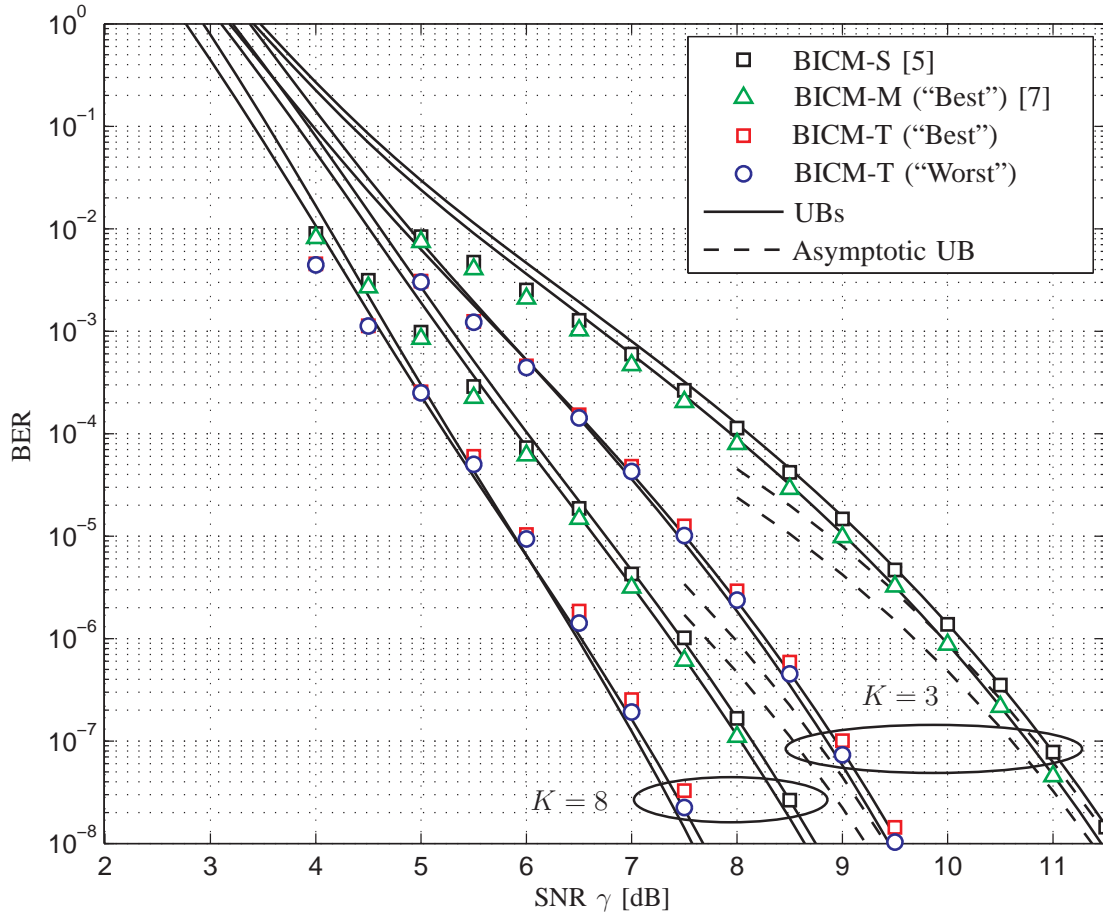


Fig. 4. BER for BICM using the (5, 7) and (247, 371) ODSCCs [14] and 4-PAM labeled with the BRGC, and for BICM-S [5], BICM-M [7], and BICM-T. The simulations are shown with markers and the UB with solid lines. The asymptotic UB is shown with dashed lines.

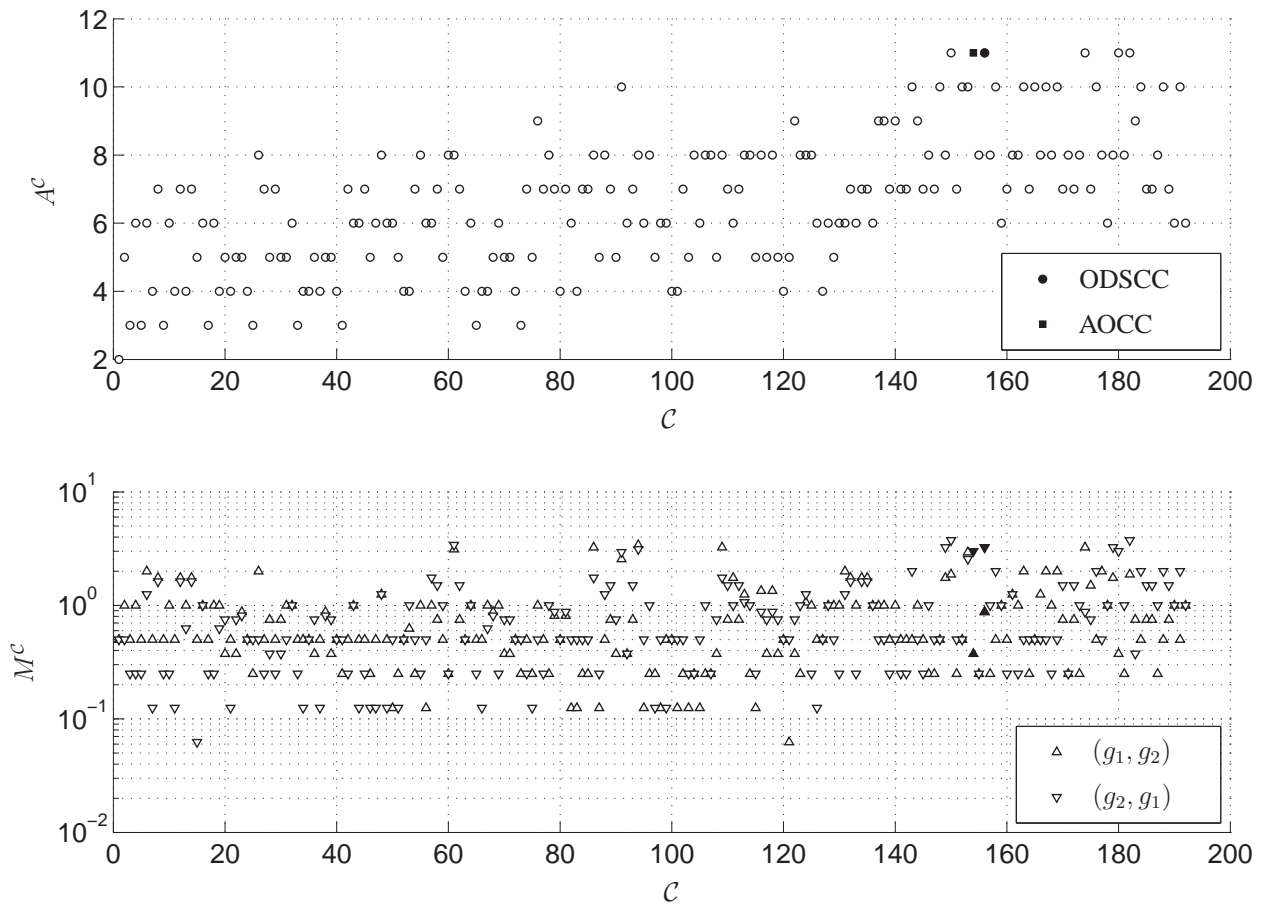


Fig. 5. Values of A^C and M^C for all the possible CCs with $K = 5$. The ODSCC and the AOCC are shown with filled markers.

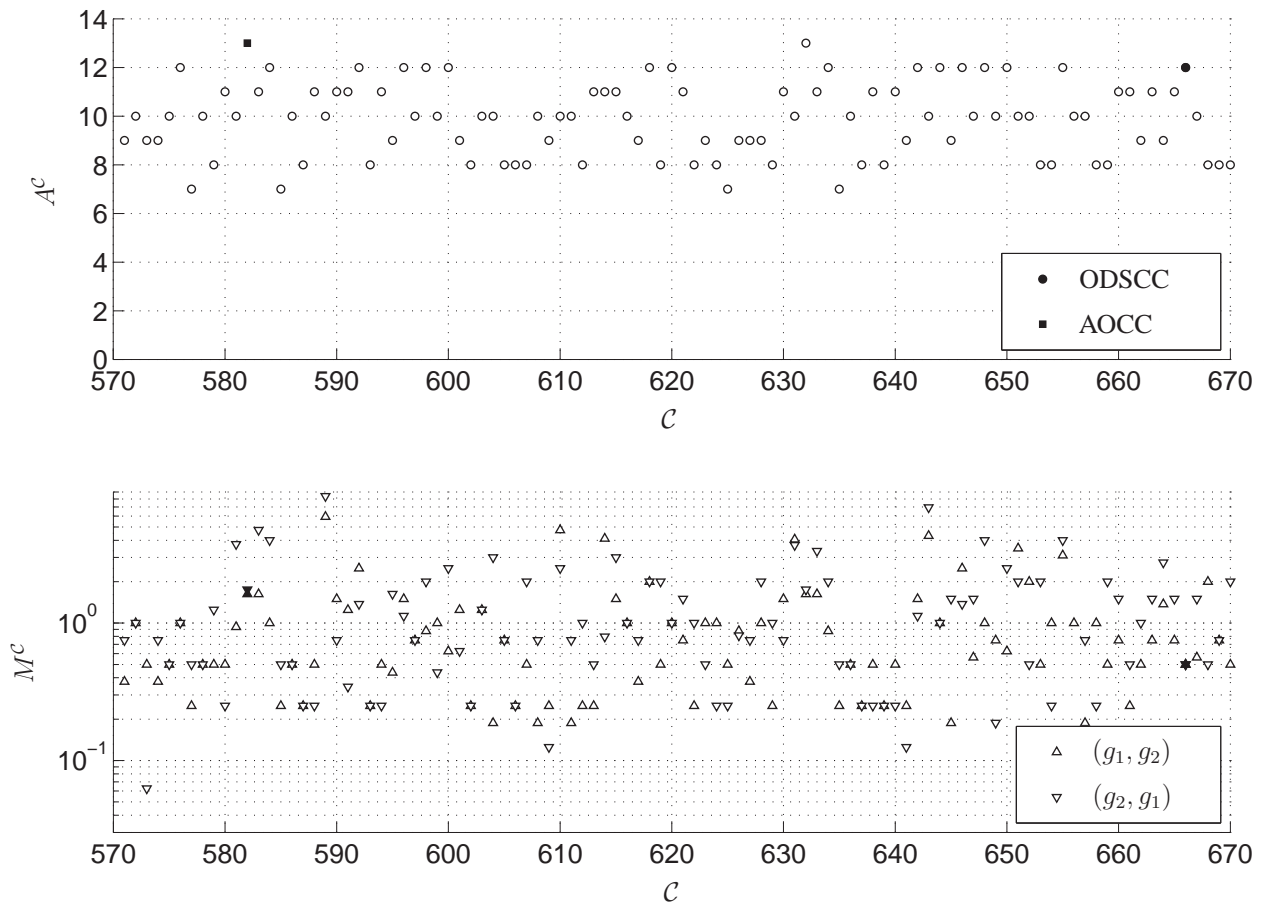


Fig. 6. Values of A^C and M^C for all the possible CCs with $K = 6$. The ODSCC and the AOCC are shown with filled markers.

TABLE I
VALUES OF $\hat{a}(\mathbf{e}_t, \mathbf{s}_t)$ AND $\hat{b}(\mathbf{e}_t, \mathbf{s}_t)$ IN (23) FOR 4-PAM FOUND BY DIRECT INSPECTION OF FIG. 3.

	$\mathbf{s}_t = [1, 1]$		$\mathbf{s}_t = [1, 0]$		$\mathbf{s}_t = [0, 0]$		$\mathbf{s}_t = [0, 1]$	
	$\hat{a}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{b}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{a}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{b}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{a}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{b}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{a}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{b}(\mathbf{e}_t, \mathbf{s}_t)$
$\mathbf{e}_t = [1, 0]$	$+\alpha/\Delta$	0	$+\alpha/\Delta$	0	$-\alpha/\Delta$	0	$-\alpha/\Delta$	0
$\mathbf{e}_t = [0, 1]$	$+\alpha/\Delta$	$+2\alpha$	$-\alpha/\Delta$	-2α	$+\alpha/\Delta$	-2α	$-\alpha/\Delta$	$+2\alpha$
$\mathbf{e}_t = [1, 1]$	$+2\alpha/\Delta$	$+2\alpha$	$+2\alpha/\Delta$	-2α	$-2\alpha/\Delta$	-2α	$-2\alpha/\Delta$	$+2\alpha$

TABLE II
VALUES OF $\hat{\mu}(\mathbf{e}_t, \mathbf{s}_t)$ AND $\hat{\sigma}^2(\mathbf{e}_t, \mathbf{s}_t)$ GIVEN IN (26) AND (27) FOR 4-PAM.

	$\mathbf{s}_t = [1, 1]$		$\mathbf{s}_t = [1, 0]$		$\mathbf{s}_t = [0, 0]$		$\mathbf{s}_t = [0, 1]$	
	$\hat{\mu}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{\sigma}^2(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{\mu}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{\sigma}^2(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{\mu}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{\sigma}^2(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{\mu}(\mathbf{e}_t, \mathbf{s}_t)$	$\hat{\sigma}^2(\mathbf{e}_t, \mathbf{s}_t)$
$\mathbf{e}_t = [1, 0]$	-3α	2α	$-\alpha$	2α	$-\alpha$	2α	-3α	2α
$\mathbf{e}_t = [0, 1]$	$-\alpha$	2α	$-\alpha$	2α	$-\alpha$	2α	$-\alpha$	2α
$\mathbf{e}_t = [1, 1]$	-4α	8α	-4α	8α	-4α	8α	-4α	8α

TABLE III
AOCCs, ODSCCs, COEFFICIENTS A^C AND M^C , AND AGs.

K	AOCCs				ODSCCs		AG [dB]	
	(g_1, g_2)	d_H^{free}	A^C	M^C	(g_1, g_2)	d_H^{free}	$\text{AG}_{\text{S} \rightarrow \text{T}}$	$\text{AG}_{\text{UC} \rightarrow \text{T}}$
3	(7, 5)	5	9	0.50	(5, 7)	5	2.55	2.55
4	(13, 17)	6	10	0.50	(15, 17)	6	2.22	3.01
5	(23, 33)	7	11	0.38	(23, 35)	7	1.96	3.42
6	(45, 55)	7	13	1.62	(53, 75)	8	2.11	4.15
7	(107, 135)	9	14	0.50	(133, 171)	10	1.46	4.47
8	(313, 235)	10	16	8.02	(247, 371)	10	2.04	5.05